Effect of R-Resonance Information on Intelligibility

Antje Heinrich & Sarah Hawkins
Department of Linguistics, University of Cambridge, Cambridge, CB3 9DA, UK

ah540@cam.ac.uk, sh110@cam.ac.uk

Abstract

We investigated the importance of phonetic information in preceding syllables for the intelligibility of minimal-pair words containing /r/ or /l/. Target words were cross-spliced into a different token of the same sentence (match) or into a sentence that was identical but originally contained the paired word (mismatch). Young and old adults heard the sentences, casually or carefully spoken, in cafeteria or 12-talker babble. Matched phonetic information in the syllable immediately before the target segment, and in earlier syllables, facilitated intelligibility of r- but not l-words. Despite hearing loss, older adults also used this phonetic information.

Index Terms: speech perception, phonetic detail, ageing.

1. Introduction

Natural speech, especially in connected discourse, is very robust and highly intelligible even in challenging listening situations such as background noise [1]. This finding may be at least partly explained by the high number of coherent linguistic cues that exist in natural speech. These cues often operate at different levels of the speech signal. For example, semantic cues operate on phrases, sentences, or paragraphs, whereas many segmental-phonetic cues tend to be restricted to one or two syllables. However, a few phonetic cues operate over a much longer time frame and affect more syllables. One such long-domain phonetic cue is r-resonance in English [2]. It arises when the following sentence base is spoken with the target word rakes: We thought we could just distinguish some rakes outside. Such r-resonance is absent when the same sentence base is spoken with the word lakes instead.

Acoustically, the presence of rakes leads to a change in formant frequencies of the vowels preceding and following [r] (used here for the English alveolar approximant). An r-resonance is typically realized as a change (often a decrease) in formant frequency in vowels preceding and following the [r]. Heid and Hawkins [3] found greatest effects in F3 and F4, with smaller ones in F2. The effect is typically strong in syllables adjacent to the [r] segment and more subtle yet long lasting in non-adjacent syllables. With some variation due to stress and segmental context, it has been measured up to around 600 ms preceding the [r] segment itself [3,4,5].

A few studies have investigated how acoustic changes in response to the presence of an /r/ affect word intelligibility. West [5] showed that the subtle acoustic changes associated with r-resonances are enough to give listeners a strong cue that an /r/ is imminent. Using minimal pairs contrasting in the presence of /r/ vs. /l/ (mirror, miller) in a carrier phrase, she showed that listeners could identify which one of the pair was in the original utterance, when the [r] or [l] itself—and sometimes more acoustic material—were replaced by noise.

In a different approach, Hawkins and colleagues [4,6] synthesized carrier sentences with or without r-resonances. Heard in cafeteria noise, the long-term resonances not only informed about the presence of the /r/ sound, but also increased intelligibility of other words in the utterances.

If r-resonance is relevant to speech perception in natural conditions, we need to show that it affects intelligibility of natural speech. The present work, part of a larger study of the use of phonetic information by elderly listeners, assesses intelligibility-in-babble of unrestricted sentences spoken in casual and careful styles. Both styles were used because it was not known which would carry most r-resonance information (its acoustic effect is more reliably measurable in unstressed syllables [3]) whereas elderly listeners would be likely to understand careful speech more easily.

The reason to assess older adults’ use of r-resonances is that they often have disproportionate difficulty understanding speech, especially consonants, in noise. However, they might use r-resonance information because it is fairly long-lasting, affects lower frequencies than many obstruent consonants, and, despite signaling the phonological function of consonants, is observable in vowels, which have relatively high amplitude. Thus there might be opportunity to hear r-resonance cues even when temporal resolution and spectral sensitivity are impaired.

2. Methods

2.1. Participants and stimulus conditions

Table 1 displays information about the participants and stimulus conditions in each of the three experiments. All volunteers were students of the University of Cambridge (young listeners), or retired residents of the city (older listeners). They were reimbursed for their time.

All participants were native speakers of British English, mostly either Standard Southern British English or Received Pronunciation (RP). Expt 1 listeners reported normal hearing. For young and old listeners in Expts 2 and 3 respectively, hearing sensitivity was measured between 250 and 8000 Hz.

<table>
<thead>
<tr>
<th>Expt</th>
<th>N</th>
<th>Speech style</th>
<th>Babble</th>
<th>s/n (dB)</th>
<th>Age range</th>
<th>Age mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41</td>
<td>Casual</td>
<td>Cafeteria</td>
<td>+2</td>
<td>17-32</td>
<td>21.4</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>Careful</td>
<td>Cafeteria</td>
<td>+2</td>
<td>18-35</td>
<td>21.2</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>Careful</td>
<td>12-talker Babble</td>
<td>+4</td>
<td>60-81</td>
<td>68.0</td>
</tr>
</tbody>
</table>

Table 1. Information about Expts. 1-3. N = number of participants per experiment; s/n = signal-to-noise ratio
2.2. Material

A 26-year-old male RP speaker read five randomized repetitions of 52 sentence pairs, like She was sitting there, just watching Terry/telly, first in a casual style, then, seven months later, more carefully but still naturally. Phoneticians judged style consistency. The speech was recorded onto a Silicon Graphics hard disk using xwaves, in a double-walled sound-treated room with a microphone about 30 cm from the speaker’s mouth. Sampling frequency was 22.05 kHz.

In each sentence pair the target word (Terry/telly) contained either one /r/ or one /l/. The rest of the sentence (the base) contained neither. Words with an [l] (l-words) were chosen as contrasts for words with an [r] (r-words) because clear [l] is thought to lack long-domain resonance [3, 7]. When possible, feet were long and target words unpredictable. Formant frequencies of some vowels were compared, and were consistent with measurements in the literature which indicate the presence/absence of r-resonances [3, 6].

To make a stimulus, a target word from one token of a sentence was spliced into the base of another token of the same sentence. The target word either matched the word it replaced (a second Terry replaced the original Terry), or mismatched it (the second Terry replaced the original telly), resulting in 104 matched and 104 mismatched stimuli. Thus, matched stimuli combined an r-base with an r-word (rr), or an l-base with an l-word (ll); mismatched stimuli combined an r-base with an l-word (rl) or an l-base with an r-word (lr).

Target phonemes were either word onset (rakes) or word medial (Terry). Word onset phonemes were in a consonant cluster (sprinter, splinter) or word initial (rakes, lakes). All target words were cut at the start of the first acoustic segment corresponding to the word, with phonetic contexts chosen to facilitate high quality splicing. Specifically, when the critical word began with an aperiodic sound, it was preceded by a periodic segment; critical words starting with periodic sounds were preceded by aperiodic or silent segments. Examples of splicing: at the boundary between periodicity offset and aperiodicity for [s] in cases like the sprinter/splinter, in the [p] silence for the prank/plank; at the boundary between aperiodicity for [s] and periodicity for /r/ in six rams/lambs. Likewise for the ends of words. Tokens chosen for splicing provided good formant continuity together with minimal differences in f0, rhythm, rate, and loudness.

For word onset /r/ or /l/, the syllable preceding the target phoneme came from a different utterance, though matched for resonance quality in half the cases. In contrast, for word medial target words, the syllable before the medial /r/ or /l/ came from the same token as the critical approximant itself, so these words’ original acoustic coherence was preserved.

After splicing, two stimulus sets were made: one masked the prank/plank; the other with British 12-talker babble. The cafeteria noise contained multiple people talking and laughing, with sounds of metal cutlery on crockery. The 12-talker babble was generated by mixing independent recordings of twelve native British English speakers of various accents reading a book or newspaper of their choice. All recordings had equal long-term RMS values to ensure that no voice stood out. Both babbles began 2 s before sentence onset: studies show that listeners adjust to background noise during the first few seconds of a presentation, which results in an increase in speech intelligibility and subsequent memory for the speech [8, 9]. The same noise segment was used for all four versions (ll, lr, rl, rr) of a sentence.

2.3. Procedure

Old listeners were not screened for audiometric loss but were expected to show loss to varying degrees. Fig. 1 shows hearing sensitivity measured at 9 frequencies to ANSI [10] standards with a GSI 61 clinical audiometer (left ear first).

All young listeners heard the speech stimuli at the same intensity level, set comfortably above the speech threshold. To ensure that old listeners heard the stimuli at a comfortable level despite their differences in hearing sensitivity, each person adjusted the intensity of a continuous speech signal to a comfortable level (mean 70.63 dB SPL). Young listeners heard the speech in cafeteria noise through Sennheiser HD 250 headphones from a PC sound card via a TEAC A-X1000 amplifier with fixed-level output. To better control sound level, old participants listened via the external input of a GSI-61 audiometer and TDH-50P headphones, and heard the speech in 12-talker British babble.

Both age groups completed a 10-sentence practice session prior to the test session to familiarize themselves with the task, the speech material and the testing procedure. The practice session was the same for all participants.

In the experiment proper, stimuli were counterbalanced across 4 groups so that each listener heard only one sentence from each set. Listeners heard the 52 test sentences in one of 4 orders, each randomly generated with the constraint that all 4 sentence types (ll, lr, rr, rl) were roughly equally distributed in the list. Sentences were 6-15 words long (mean: 10.3). Listeners heard each sentence while seeing its first 1-5 words (mean 2.3) in an answer window, then typed the rest of the words in the window. On average, the target word was the 6th to be typed. The task was self-paced and took 20-30 minutes. Testing was in a double-walled sound-attenuated room.

3. Results

Total Ns were adjusted to account for the few (0.2%) missing data points. All results are presented as percent correct target words. For Expt. 1, performance was also calculated based on correct target phoneme (/r/ or /l/) rather than target word. Consistent with previous work [6], it yielded the same pattern of results though with about 20% better performance.

Responses were analyzed in two ways to answer the following questions:

1) Does the presence of a congruent (matched) base improve the intelligibility of a target word that follows?

2) Is congruent information in the syllable immediately prior to the target word especially important for intelligibility?

In all analyses, percent correct identified words for each of the four splicing conditions, matched (ll, rr) and mismatched (lr, rl), is shown separately according to whether the target phoneme was (a) in the word onset or (b) word medial.
3.1. The importance of congruent phonetic cues in the preceding sentence base

Figure 2 shows percent correct target words for all three experiments, distinguishing word onset and word medial stimuli. Table 2 shows repeated-measures ANOVA results (2 sentence base (l/r) x 2 match (matched/mismatched)). For word onset targets, r-bases did not help intelligibility. For r-bases with word-medial targets, all three experiments have similar patterns; matched stimuli were more intelligible (rr > rl). In contrast, for l-bases with word-medial targets, matching (l) does not facilitate the correct perception of the target word in Expts 1 and 2 (l = lr). In Expt 3, the small ll > lr difference only contributes to the main effect of match.

![Figure 2. Mean percent correct for target words for four splicing conditions (ll, lr, rr, rl) in Expts 1-3. Black squares: target phoneme is word onset. Open triangle: target phoneme is word medial. Error bars: the standard error of the mean after between-subject variability has been removed [11].]

Next, we asked whether the effect of r-resonance is specific to a particular speech style (Expts 1 vs 2). For word-onset phonemes, careful speech was about 25% more intelligible than casual speech (panel 2 vs. panel 1 in Fig. 1: F[1,63] = 85.07, p < 0.001, in a 2 base (l/r) x 2 match (matched/ mismatched) x 2 style (casual/careful) mixed-model ANOVA with base and match as within- and speech style as between-subject variables). There were no other significant effects in this ANOVA on word-onset targets. Word-medial targets had a more complex pattern. Careful speech was again more intelligible than casual (F[1,63] = 95.06, p < 0.001), but both base x match and base x match x style interactions were also significant (respectively F[1,63] = 73.17, p < 0.001; F[1,63] = 5.24, p = 0.025). Posthoc t-tests confirm that in both interactions, l = lr and rr > rl, with the rr vs rl difference greater for careful speech. Thus, in these word-medial conditions, r-resonances strongly affect word intelligibility, especially in careful speech.

To examine age effects, we ran mixed-model ANOVAs (age x l/r base x match, with age group as between- and base and match as within-subject variables) on Expts 2 and 3 for word-onset and word medial phoneme location conditions separately. For word onset targets, there was no effect of age. As intended, this reflects the more favourable s/n ratio used for older listeners (+4 dB rather than ~2 dB). The only significant effect was an interaction between base and match (F[1,50] = 8.68, p = 0.005; ll = lr, rr > rl) reconfirming that for r- but not l-bases, matching base with target word increased intelligibility.

However, when the target phoneme was word medial in Expts 2 and 3, age, match, base x match, and base x match x age were all significant. The base x match interaction is as in the previous analyses: ll = lr, rr > rl (F[1,50] = 27.36, p < 0.001). The three-way interaction (F[1,50] = 9.52, p = 0.003) shows that matching r-base with word-medial r-targets affects intelligibility more for young than old listeners: rr > rl in both groups, but the difference is greater for young people. Other main effects were age (young>old, F[1,50] = 5.12, p = 0.028) and match (match > mismatch, F[1,50] = 6.02, p = 0.018).

Overall, then, for word-medial stimuli, the age groups differ in that older adults benefit from matched stimuli, both ll and rr, whereas young adults only benefit from the rr condition, but to a greater degree than the older listeners. Future work will address to what extent this difference is dependent on age vs. type of babble, but it may be that, in these listening conditions, older adults rely on the acoustic signal more than young adults do.

### Table 2. Repeated-measures ANOVA results for base (l/r) x match (match/mismatch) for Experiments 1, 2 and 3. See text for details.

<table>
<thead>
<tr>
<th>Expt</th>
<th>Phoneme location</th>
<th>Effect</th>
<th>F (df)</th>
<th>P</th>
<th>Posthoc t-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Onset</td>
<td>No significant difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medial Base x Match</td>
<td>30.9 (1, 40)</td>
<td>0.001</td>
<td>ll = lr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rr &gt; rl</td>
</tr>
<tr>
<td>2</td>
<td>Onset</td>
<td>No significant difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medial Base x Match</td>
<td>37.4 (1, 23)</td>
<td>0.001</td>
<td>ll = lr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rr &gt; rl</td>
</tr>
<tr>
<td>3</td>
<td>Onset</td>
<td>No significant difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medial Base x Match</td>
<td>9.61 (1, 27)</td>
<td>0.004</td>
<td>ll = lr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rr &lt; rl</td>
</tr>
<tr>
<td></td>
<td>Medial Match</td>
<td>4.04 (1, 27)</td>
<td>0.054</td>
<td>Match &gt; Mismatch</td>
<td></td>
</tr>
</tbody>
</table>

To examine the effect of hearing loss on the intelligibility results for older listeners, we entered hearing sensitivity thresholds as a covariate into the ANOVA. Hearing loss did not systematically covary with the intelligibility results.

3.2. The importance for intelligibility of resonance information in the syllable immediately prior to the target segment

We have already shown that the presence of congruent resonance information across the whole sentence base facilitates identification of r-words in noise for both age groups. To examine whether congruent resonance cues in the syllable immediately before the target segment disproportionately facilitate word intelligibility, we compared word onset and word medial phoneme location conditions with the match/mismatch variable pooled (l-words (l) = ll + rl; r-words (r) = lr + rr).

Recall that what distinguished these conditions, other than l- and r-words, was whether the syllable immediately before the target phoneme, /l/ or /r/, came from the same or a different utterance token. The syllable before a word onset /l/ or /r/ came from a different utterance token, though in half the cases it was matched for resonance quality. The syllable before a word medial /l/ or /r/ came from the same token.

Table 3 summarizes the results of the 2 x 2 repeated-measures ANOVAs (target word (l/r) x phoneme location (word onset/word medial)) that show that phoneme location affected word intelligibility in all three experiments. In Expt 1 (casual speech), the effect of phoneme location was significant both as a main effect (words with medial /l/ or /r/ were more intelligible than words with onset /l/ or /r/) and in interaction with r-words vs. l-words. In Expts 2 and 3, the effect of phoneme location was only expressed in the word x location
interaction. This interaction between r- and l-words and phoneme location was similar across all three experiments insofar as a medial phoneme location improved intelligibility of r-words. More precisely, the factor that increased intelligibility was the acoustic integrity of words containing /r/. The acoustic integrity of l-words was either inconsequential (Expt 1), or else word onset /l/ facilitated intelligibility compared with word medial /l/ (Expts 2 and 3).

<table>
<thead>
<tr>
<th>Expt</th>
<th>Phoneme location</th>
<th>F (df)</th>
<th>p</th>
<th>Posthoc t-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Word x phoneme location</td>
<td>9.25 (1, 40)</td>
<td>0.004</td>
<td>Word-medial &gt; word-onset</td>
</tr>
<tr>
<td>2</td>
<td>Word x phoneme location</td>
<td>15.64 (1, 40)</td>
<td>0.001</td>
<td>l&lt;sub&gt;o&lt;/sub&gt; = l&lt;sub&gt;b&lt;/sub&gt; r&lt;sub&gt;o&lt;/sub&gt; &gt; r&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
<tr>
<td>3</td>
<td>Word x phoneme location</td>
<td>9.65 (1, 27)</td>
<td>0.004</td>
<td>l&lt;sub&gt;o&lt;/sub&gt; &lt; l&lt;sub&gt;b&lt;/sub&gt; r&lt;sub&gt;o&lt;/sub&gt; &gt; r&lt;sub&gt;b&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Table 3. Statistical results for 2 word (r/l) x 2 phoneme location (word onset/word medial) repeated-measures ANOVAs for all three experiments. o: word onset; m: word medial.

4. Discussion

Target words containing an /r/ or /l/ were cross-spliced to investigate the effect of long-domain phonetic cues on their intelligibility in read sentences.

For r-words, the presence of congruent acoustic-phonetic information in preceding syllables facilitated intelligibility of the target word in all but one analysis (Expt 3, older listeners in word-onset conditions). The same was not true for l-words. There, congruency of preceding information was mostly irrelevant. The older listeners tested were also influenced by r-resonances, independent of the mild hearing loss they all possessed to some degree. Presumably some correlation with hearing loss would appear in a sample exhibiting a greater degree and range of loss than the present sample.

The perceptual effect of r-coloring on the syllable immediately preceding the /r/ is interesting. In addition to the general long-domain effect of r-resonance on intelligibility, r-resonance in the syllable directly before the target phoneme affected target word intelligibility irrespective of the presence or absence of congruent information in the preceding signal. When the original r-words were intact, they were especially intelligible. In contrast, intelligibility of l-words was not affected by whether the original word was intact; rather, hearing an /l/ at the beginning of the word facilitated its intelligibility most. This suggests that the strong effect of r-coloring measurable in the immediately preceding syllable [3] is especially salient perceptually, and that any sort of splicing, even to a matched word, can reduce its salience. This result demonstrates the importance to perception of acoustic continuity at quite a fine degree of detail.

In summary, the intelligibility of words containing an /r/ is significantly increased by the presence of resonances due to that /r/ in the immediately preceding syllable and, to a lesser but still significant extent, over several earlier syllables. These principles are now confirmed for several types of speech. They have already been shown for synthetic speech [4,6] (see [4] for the spread of resonances after an /r/) and for words in a single carrier phrase in natural speech [5]. The present study is the first to extend these findings to natural speech, namely meaningful single sentences with unrestricted text, read fluently in both casual and reasonably clear/careful but still natural speaking styles. The implication is that listeners benefit from the presence of r-resonances during speech processing. This is consistent with conclusions drawn in the literature, but was by no means an inevitable outcome.

To build on previous work, Expts 2 and 3 confounded listeners’ age with type of babble. Cafeteria noise was needed in Expts 1 and 2 to compare our naturally-spoken unrestricted text with previous intelligibility tests on more restricted synthetic speech. But cafeteria noise is less predictable than 12-talker babble. Having established that young listeners use r-resonances to understand careful and casual read speech in cafeteria noise, and that older listeners use them in careful speech in 12-talker babble, future work will examine whether the type of babble affects intelligibility, and thus to what to attribute the small observed differences between age groups.

Of special note is that the effect of r-resonances on intelligibility does not appear to be seriously affected by the hearing loss that is normal with aging. These results are excellent news for older listeners because they demonstrate another linguistic source of information, besides semantic cues [12], that older adults have at their disposal to compensate for the consequences of hearing loss. This case is especially interesting because it is rooted in subtle, long-lasting spectral cues, which give very weak phonological/semantic support since they only cue a particular phoneme rather than a word or phrase dictated by context.

5. Acknowledgements

We thank ERA-AGE (FLARE) for funding Antje Heinrich, and Yvonne Flory and Pia Rübig for helping with stimulus preparation and data collection.

6. References